

Who is bragging more online? A large scale analysis of bragging in social media



Overview

- Bragging is a speech act that involves disclosing **positive content** about oneself or their close network
- Bragging behavior was quantitatively studied in specific contexts through **manual analysis of small data sets** of hundreds of posts
- We conduct the **first large scale quantitative study** of bragging on social media, focusing on understanding the **prevalence, temporal dynamics** and **user factors** impacting bragging prevalence
- We use a longitudinal data set including over **1 million English tweets** posted by a **group of 2,685 Twitter users** (U.S.A) over 10 years.
- We conduct an extensive **linguistic analysis** to unveil **specific bragging themes** associated with different time period and user traits

Example Tweets

Type	Text
Bragging	<i>Just impressed myself with how much French I think I understood! One semester at KC FTW!</i>
Non-bragging	<i>Glad to hear that! Well done Jim!</i>

Measuring Bragging Prevalence

- Data for Model Training** - We use a bragging data set consisting of **6,696 English tweets** and each tweet is manually annotated as bragging (781, 11.66%) and non-bragging (5,915, 88.34%)
- Predictive Model** - We re-implement the best performing predictive model, **BERTweet-LIWC**, in Jin et al. (2022) on identifying whether a tweet contains bragging or not (72.51 macro F1)

Bragging Prevalence Metrics

- Overall Bragging Prevalence** - We compute the distribution of average bragging percentage across users for each month:

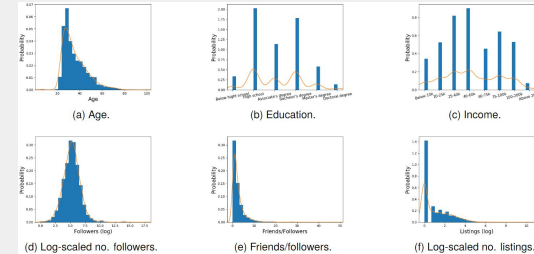
$$A(a_1, \dots, a_n) = \frac{\sum_{i=1}^n a_i}{n} \quad (n: \text{number of months})$$

- User Bragging Prevalence** - 1. We obtain a time-normalized bragging distribution for each user in each time window: $p_i(a_1, \dots, a_n) = \frac{\sum_{t=1}^n a_{i,t}}{\sum_{t=1}^n 1}$
- 2. We average the normalized bragging distribution for each user:

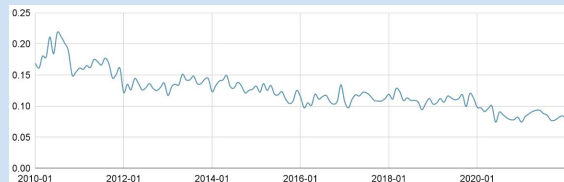
$$l = \frac{\sum_{i=1}^n p_i}{n}$$

Analysis Data

- A Set of Twitter Users** - 1. We combine three data sets where users associated with socio-demographic characteristics; 2. We collect all historical tweets from these users
- Data Filtering** - We filter out non-English and duplicate tweets, replies and retweets, tweets that are automatically generated from third parties and users that have posted fewer than 20 tweets. The analysis data contains 2,685 users in total
- Computing Bragging Ratio** - 1. We use the predictive model to identify bragging or not of all 1,031,276 tweets; 2. Then we compute the normalized bragging ratio for each user using bragging prevalence metrics
- User Demographic Traits** - Gender, Age, Education, Income
- User Popularity Traits** - No. Followers, No. Friends/No. Followers, No. Listings



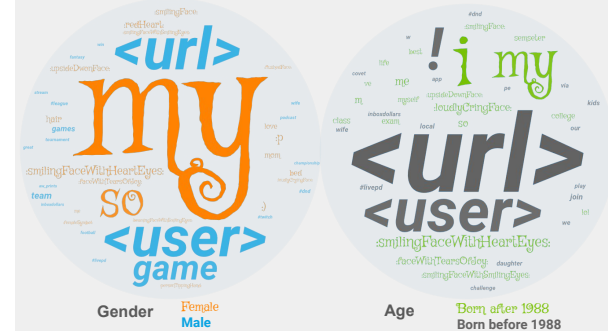
Bragging Prevalence over Time



Demographic and Popularity Factors

Trait	r	p_{unc}	p_{corr}
User Demographics			
Gender ($\varrho-1, \sigma^2=0$)	0.10	<.001	<.001
Age	-0.16	<.001	<.001
Education	0.14	<.001	<.001
Income	0.07	<.003	<.002
User Popularity			
No. Followers	0.12	<.001	<.001
No. Friends/Followers	-0.10	<.001	<.001
No. Listings	0.09	<.001	<.001

Feature Correlations



Please check out our paper for more detailed analysis!

Findings

- Bragging steadily **decreases** over time in the past 10 years within the same group of users
- Bragging is more prevalent among users who are **female, generally younger, more educated**, have a **higher income** and more popular on Twitter
- Male** users and users with **higher income** brag more about **leisure activities**; **Female** users and users with **lower education** focus more on **themselves** when bragging; Bragging by **older users** and users who have **higher education** are more likely to **involve others**; **Emojis** are more frequently used by **female** and **younger** users while bragging